# AMD Athlon™ Processor and AMD Duron™ Processor with Full-Speed On-Die L2 Cache

## Enabling an Innovative Cache Architecture for Personal Computing

ADVANCED MICRO DEVICES, INC.
**One AMD Place**
**Sunnyvale, CA 94088**

## Introduction: The AMD Athlon™ Processor with Full-Speed On-Die L2 Cache

At its introduction, the AMD Athlon™ processor marked the arrival of the world's first seventh-generation microarchitecture and in doing so set a new performance standard for x86 processors. Among the processor's award-winning architectural features are a revolutionary 200MHz, 1.6Gbytes/sec system bus, a fully pipelined, superscalar floating point engine, and an enhanced version of AMD's 3DNow!™ technology. Additionally, the processor offers 128KB of L1 cache—four times the L1 cache of competing x86 processors—along with a 512KB external backside L2 cache running at up to half the speed of the processor core.

With new versions of the AMD Athlon processor, however, AMD has improved overall system performance by integrating the processor's L2 cache directly onto the processor die. This white paper describes the benefits of utilizing the AMD Athlon processor with *full-speed on-die L2 cache*.

The new AMD Athlon processor boasts three times the full-speed on-die cache of previous AMD Athlon processors. The new processor features 128KB of L1 cache—plus 256KB of full-speed, on-die L2 cache—for a total internal system cache of 384KB.

On-die L2 cache allows the L2 cache speed to scale with the processor speed one-to-one, enabling a more substantial level of performance across a broad range of memory-intensive applications. Additionally, integration of the L2 cache onto the die eliminates the need for external, high-performance L2 cache SRAM, enabling the PGA form factor.

With the AMD Athlon processor's migration to 0.18-micron process technology completed, the processor's integrated cache design does not significantly change the size of the die, maintaining ease of manufacture. The advanced 0.18-micron manufacturing process enables smaller transistor sizes in comparison to the 0.25-micron process. Therefore, the space penalty for the integration of the L2 cache is minimal, as the L2 cache accounts for only 20% of the entire processor die. The new AMD Athlon processor's 37-million-transistor-die is ~120 $mm^2$ as compared to the ~102 $mm^2$ die size for the original AMD Athlon processor.

## The Benefits of "Exclusive" Cache Architecture

The new AMD Athlon processor with full-speed on-die L2 cache features an *exclusive* cache architecture as opposed to the *inclusive* cache architecture utilized by previous-generation x86 processors.

An inclusive cache architecture requires the L2 cache to duplicate every cache block held by the processor's L1 cache. More specifically, for every cache block in an inclusive cache architecture's L1 cache, the L2 cache must contain the same redundant data, thereby decreasing the effective L2 cache available for new information. Therefore, assuming an L1 cache of 32KB, that processor's inclusive L2 cache size is effectively 32KB smaller than its given physical size.

In contrast, with an exclusive cache architecture, the L2 cache contains only *victim* or *copy-back* cache blocks that are to be written back to the memory subsystem as a result of a conflict miss. These terms, *victim* or *copy-back*, refer to cache blocks that were previously held in the L1 cache but had to be overwritten (evicted) to make room for newer data. On the new AMD Athlon processor, this exclusive cache architecture enables a full 256KB of L2 cache and 128KB of L1 cache for a total dedicated storage space of 384KB.

Regardless of cache architecture, the primary function of the L1 cache is to store frequently used data that can be easily accessed by the processor. This enables a high cache *hit rate*, defined as the percentage of time that requested data will be found in the cache instead of main memory. The larger the L1 cache, the greater the advantage to overall processor performance. Should the L1 cache become filled, the evicted L1 victim or copy-back cache block is simply moved to the L2 cache. This keeps the once often-used data close to the processor as opposed to having it moved out to main memory. At the same time, because of its exclusive architecture, the L2 cache does not waste space by duplicating the contents of the L1 cache.

## Improved L2 Cache Efficiency through Lower Latency, Increased Bandwidth and Greater Associativity

### *Latency*

When compared to previous-generation x86 processors, the new AMD Athlon processor's internal cache, with its large L1 cache and exclusive L2 cache architecture, is more likely to contain the requested data. Data found in a cache can be used sooner than data in main memory. Therefore, cache is said to have lower latency. With more data in the cache, overall system *latency*—the time required to access the next piece of critically needed memory—is lower. Consequently, overall system performance is higher.

Integrating the L2 cache onto the processor die significantly lowers hit latency, as it takes much less time to move data across the die than is required to read it from an external SRAM. The lower the latency, the shorter the response-time required for receipt of the

requested data. The AMD Athlon processor with full-speed on-die L2 cache delivers more than 45% lower latency compared to previous AMD Athlon processors. This low latency is due in part to the on-die nature of the cache memory as well as the processor's exclusive L2 cache architecture—an architecture not found in the original AMD Athlon processor or previous-generation x86 processors.

The average L2 cache latency of the new AMD Athlon processor often cannot be properly determined using today's common cache-testing utilities. These types of cache utilities typically measure L2 latencies in terms of worst-case delays, which are rarely encountered in real-world applications. Additionally, due to the exclusive cache architecture of the new AMD Athlon processor, the testing methodology of these utilities must be altered in order to obtain an accurate L2 latency average.

The new AMD Athlon processor's *victim buffer* contains data evicted from the L1 cache. It features up to eight 64-byte entries, where each entry corresponds to a 64-byte cache line. It's important to note that the new AMD Athlon processor's L1 and L2 caches work with 64-byte cache lines, equal to twice the Pentium III processor's 32-byte cache line capability.

Due to its large (128 KB) capacity, the AMD Athlon processor's L1 cache is capable of efficiently handling most requests for data. As a result, the victim buffer can be drained during idle cycles to the L2 cache interface. In real-world applications, the new AMD Athlon processor's victim buffer is rarely full. Consequently, the L2 load-use latency will only be 11 cycles—which includes a 3 cycle L1 miss access. The actual time it takes for the L2 cache to provide the first critical word back after it receives a request is 8 cycles, as shown in *Figure 1* below.

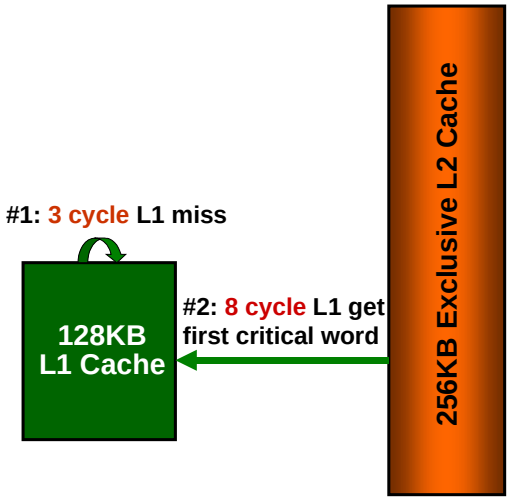


**Figure 1: AMD Athlon™ Processor "Victim Buffer Not Full"**

*Figure 2* below shows the latency of each step in an L1 miss/L2 hit scenario when the victim buffer is full. When the victim buffer reaches capacity and there is an L1 miss/L2 hit, the victim data is then copied from the victim buffer back to the L2 cache—resulting in the *worst-case* delay scenario that older cache-testing utilities typically report.

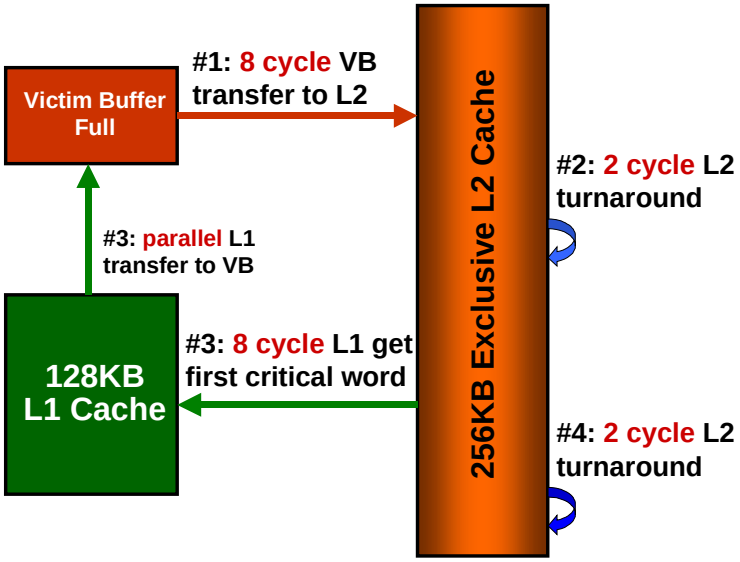


**Figure 2: AMD Athlon™ Processor "Victim Buffer Full"**

Properly measured, the true latency of the new AMD Athlon processor is between 11 and 20 clock cycles, depending on processor activity. Under real-world application workloads, the average L2 cache latency of the new AMD Athlon processor rests at approximately 11 clock cycles.

The new AMD Athlon processor's large L1 cache and exclusive L2 cache architecture contribute to the processor's clock-for-clock performance advantage over Intel's Pentium® III processor.[1]

### *Cache Bandwidth*

The increase in cache speed enabled by the integration of the L2 cache onto the processor die increases L2 cache bandwidth by 300% over previous AMD Athlon processors. Higher bandwidths allow the processor to work on more data over the same

[1] For specific benchmark results, visit http://www.amd.com/products/cpg/athlon/benchmarks/benchmarks.html

amount of time. For example, the previous 1000MHz AMD Athlon processor requires 24ns (3 clocks x 8 transfers) to transfer a 64-byte cache line while the new AMD Athlon processor only requires 8ns.

Unlike competing x86 processors—including Intel's Pentium III processor—the AMD Athlon processor features an L1 cache large enough to enable a high hit rate while significantly reducing and minimizing bandwidth demands on the processor's L2 cache. For this reason, expanding the width of the L2 bus—currently 64 bits—offers little or no benefit to overall performance. Emphasizing the high cache hit rates enabled by the new AMD Athlon processor's large L1 cache and exclusive L2 cache architecture, extensive application performance testing shows that increasing the L2 bus width from 64 bits to 256 bits offers little or no significant impact to overall processor performance. Processors with much smaller L1 caches need to be refilled more often and typically require a higher L2 bandwidth interface to help offset the shortcomings of their small L1 cache size.

### *Set Associativity*

The new AMD Athlon processor's 16-way set associative cache is eight times more associative than previous AMD Athlon processors which feature a 2-way set associative cache. Increasing the set associativity increases the hit rate by reducing data conflicts. This translates into more possible locations in which important data can reside without having to throw away other often-used data in the L2 cache. If the requested data is not found in cache memory, the processor must search for it in main memory. Because main memory is limited by its lower speeds—currently no faster than 133MHz—cache hits significantly improve the average memory access. The more effective the cache, the better the average memory performance, the more improved the overall application performance.

## Fill Buffers, Bus Queue Entries and Write-Back Buffers

The new AMD Athlon processor, like previous AMD Athlon processors, includes eight fill buffers, eight bus queue entries, and eight write-back buffers. Intel's Pentium III processor only has six fill buffers, eight bus queue entries, and four write-back buffers. The more buffers that are dedicated to holding data for the microprocessor to process, the less likely the processor will stall while waiting for data to be delivered to the processor. Likewise, the more buffers that are available the more likely it is that data can be assigned to a buffer for execution, thereby increasing data-processing speed. These architectural features contribute to the AMD Athlon processor's ability to execute near the peak bus bandwidth of 1.6 GB/s.

## The AMD Duron™ Processor

This white paper details the features and benefits specific to the new AMD Athlon processor. It should be noted, however, that with the exception of a smaller 64KB L2 cache, the AMD Duron™ processor's cache architecture is identical to that of the new AMD Athlon processor with full-speed on-die L2 cache.

## Summary: Enabling Improved Performance on Memory-Intensive Applications

Featuring the world's first seventh-generation microarchitecture, the new AMD Athlon processor is among the most powerful x86 processors for high-performance desktop PCs.

With its L2 cache integrated onto the die, AMD Athlon processor performance is taken a step further. Through its exclusive L2 low-latency cache architecture and high associativity, the new AMD Athlon processor enables an improved level of performance on a broad range of memory-intensive applications.

These applications include business and personal productivity suites, multimedia and entertainment software, and still/video image-manipulation applications, with the most significant performance gains occurring in high-end CAD/CAM applications. These wide-ranging performance improvements demonstrate AMD's continuing ability to provide reliable, high-performance processing solutions for both consumer desktop and enterprise computing systems.

## AMD Overview

AMD (NYSE: AMD) is a global supplier of integrated circuits for the personal and networked computer and communications markets. AMD produces processors, flash memories, and products for communications and networking applications. The world's second-leading supplier of Windows® compatible processors, AMD has shipped more than 120 million x86 microprocessors, including more than 90 million Windows compatible CPUs. Founded in 1969 and based in Sunnyvale, California, AMD has sales and marketing offices worldwide and manufacturing facilities in Sunnyvale; Austin, Texas; Dresden, Germany; Bangkok, Thailand; Penang, Malaysia; Singapore; and Aizu-Wakamatsu, Japan. AMD had revenues of $2.8 billion in 1999.

## Cautionary Statement

This release contains forward-looking statements, which are made pursuant to the safe harbor provisions of the Private Securities Litigation Reform Act of 1995. Forward-looking statements are generally preceded by words such as "expects," "plans," "believes," "anticipates," or "intends." Investors are cautioned that all forward-looking statements in this document involve risks and uncertainty that could cause actual results to differ materially from current expectations. Forward-looking statements in this document about the AMD Athlon processor involve the risk that the AMD Athlon system bus will not support the requirements of next-generation system platforms; that AMD may not be successful in developing an infrastructure to support the processor; that third parties may not provide peripherals or the infrastructure to support the processor and the processor's system bus; and that the processor will not achieve customer and market acceptance. We urge investors to review in detail the risks and uncertainties in the company's Securities and Exchange Commission filings, including the most recently filed Form-10K.